

BUNDESREPUBLIK DEUTSCHLAND

DE 00/442

248

133

**PRIORITY
DOCUMENT**SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 14 AUG 2000

WIPO

PCT

09/913615 4

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung****Aktenzeichen:**

199 48 546.1

Anmeldetag:

8. Oktober 1999

Anmelder/Inhaber:

VOXAR AG, München/DE

Erstanmelder: Voxar AG in Gründung,
München/DE**Bezeichnung:**Verfahren und Vorrichtung zur darstellungsoptimier-
ten Übertragung von Video- und Audiodaten**IPC:**

H 04 N 7/14

**Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Patentanmeldung.****München, den 2. August 2000
Deutsches Patent- und Markenamt
Der Präsident**

Im Auftrag

Weihmayer

VOXAR AG
VOX/P991

08.10.99



**Verfahren und Vorrichtung zur darstellungsopti-
mierten Übertragung von Video- und Audiodaten**

5

Priorität: DE-A-19906 472.5 vom 16. Februar 1999

10

Stand der Technik

15

Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur darstellungsoptimierten Übertragung von Video- und Audiodaten gemäß den Oberbegriffen der Patentansprüche 1 und 15.

20

Es gibt bereits Kommunikationsdienste und -systeme, die neben dem auditiven Bereich auch visuelle Medien bzw. Kanäle zur Übermittlung von Ton- und Bildinformation anbieten.

25

Bereits heute kommen z. B. bei Videokonferenzsystemen Bildbearbeitungsalgorithmen zum Einsatz (WO96/09722).

30

Die DE-0402895 C1 beschreibt ein Verfahren und eine Vorrichtung zur Korrektur des Augenwinkels bei Arbeitsplatzvideosystemen. Hier wird der Augenbereich eines Anwenders editiert, um den Eindruck zu erwecken, man sehe direkt in die Kamera statt auf den Monitor. Diese

08.10.99

5

2

Editierfunktion arbeitet unbemerkt vom Sender und Empfänger der Übermittlung.

Es ist weiterhin bekannt, dass es im Fernsehbereich
5 Verfahren gibt, die Informationen vor der Weitergabe an
die Sendeeinheit in Echtzeit oder fast in Echtzeit zu
verändern. Vorteil dabei ist, dass die geänderten
Informationen der vom Ersteller gewünschten Darstellung
entsprechen bzw. nahe kommen. Zu diesen Verfahren gehört
10 "Blue Box", "Maskierung" und "Automaskierung".

Wie bekannt, übertragen die oben genannten audiovisuellen
Kommunikationsdienste und -systeme mehr Informationen als
das Telefon. Zu diesen Informationen zählen unter anderem
15 das Erscheinungsbild des sendenden Teilnehmers, seine
Körpersprache, Mimik und Gestik sowie die Umgebung, die
von der Videokamera eingefangen werden kann. Dabei
entsteht jedoch das Problem, dass die Übertragung dieser
visuellen Informationen zu psychologischen Hemmschwellen
20 bei der Nutzung der oben genannten audiovisuellen
Kommunikationsdienste und -systeme führt. Es gilt als
sicher, dass diese psychologischen Hemmschwellen zum
großen Teil dazu beigetragen haben, dass sich
beispielsweise das Bildtelefon immer noch nicht auf dem
25 Markt durchgesetzt hat.

Aufgabe und Vorteile der Erfindung

Der vorliegenden Erfindung liegt die Aufgabe zugrunde,
30 ein Verfahren und eine Vorrichtung zu ermöglichen, die
Anwendern der genannten audiovisuellen

Kommunikationsdienste und -systeme die Kontrolle über den Austausch selbstbezogener Informationen ermöglicht.

5 Diese Aufgabe wird durch ein Verfahren und eine Vorrichtung jeweils mit den Merkmalen und Mitteln der Ansprüche 1 und 15 gelöst.

10 Dem sendenden Teilnehmer von audiovisuellen Kommunikationssignalen wird durch die Erfindung die Möglichkeit gegeben, den Inhalt so zu beeinflussen, dass der gesendete Inhalt der Vorstellung des sendenden Teilnehmers über seine Darstellung nahekommt bzw. mit seiner Vorstellung übereinstimmt. Damit wird folgender Nachteil der bisherigen audiovisuellen Kommunikation überwunden:

15 Der sendende Teilnehmer einer audiovisuellen Nachricht kann unabhängig von seiner derzeitigen Erscheinung, z. B. morgens nach dem Aufstehen, bei Krankheit, Missgestaltungen des Gesichts oder trotz eines aus anderen Gründen unvorteilhaften Erscheinungsbildes ungehindert kommunizieren.

20

Durch die Verwendung der vorliegenden Erfindung, z. B. bei der Bildtelefonie, bei Videokonferenzen, Arbeitsplatzkonferenzen, Internet-Konferenzen, etc. stellt ein

25 jeweiliger Teilnehmer an einer audiovisuellen Kommunikation sicher, dass sein Erscheinungsbild und die von der Kamera ebenfalls eingefangene Umgebung beim empfangenden Teilnehmer seinem inneren Selbstwertverständnis entspricht. Somit schützt die Verwendung der vorliegenden

30 Erfindung die Privat- und Intimsphäre.

Gemäß einem wesentlichen Aspekt der Erfindung ist ein die obige Aufgabe lösendes Verfahren zur darstellungs- optimierten Übertragung von Videodaten und/oder Audio- daten, insbesondere beim Bildfernsprechen, bei dem von
5 einer Video- und Audioquelle kommende Bild- und Audio- daten vor ihrer Übertragung zu einem Empfänger anhand von Referenzinformation entsprechend wenigstens einem vorbe- stimmten oder vorbestimmbaren Kriterium verändert werden, dadurch gekennzeichnet, dass das Verfahren wenigstens
10 folgende Schritte aufweist:

- a) Referenzbildinformation wird zuvor erstellt und gespeichert;
- b) Parameter, die Bildinformation betreffen, werden zuvor definiert, gespeichert und der in Schritt a)
15 gespeicherten Referenzbildinformation zugeordnet;
- c) ein von einer Videoquelle kommendes Bild wird hinsichtlich eines oder mehrerer ausgewählter (auswähl- barer) Bildinformationsparameter von den in Schritt b) definierten und gespeicherten Parameter extrahiert;
- 20 d) das Bild von Schritt c) wird basierend auf dem oder den ausgewählten Bildinformationsparameter(n) anhand der zugeordneten Referenzbildinformation editiert; und
- e) das in Schritt d) editierte Bild wird zu einem oder mehreren Empfängern übertragen.

25

Die Schritte d) und e) können bei einer Ausführungsart an einem zentralen Ort ausgeführt werden, der vom sendenden Teilnehmer entfernt ist, wobei das Bild, das zugeordnete Referenzbild und der oder die Bildinformationsparameter
30 vom Ort des sendenden Anwenders zum zentralen Ort übertragen werden.

08.10.99

8

5

Alternativ können die Schritte a) bis e) am Ort eines sendenden Anwenders ausgeführt werden.

5 Um sicher zu stellen, dass nur ein berechtigter Anwender die anwenderspezifische Optimierung durchführen und mit editierten Bildern und/oder veränderter Stimmlage auf Sendung gehen kann, weist das Verfahren einen zusätzlichen Schritt f) auf, mit dem zuvor gespeicherte Sprachproben berechtigter Anwender mit einer gesprochen Kennphrase eines aktuell sendenden Anwenders verglichen werden und bei positivem Vergleichsergebnis eine Sende-
10 freigabe erfolgt.

15 Dabei wird bevorzugt im Schritt f) die gespeicherte Sprachprobe einem Referenzbild zugeordnet, das zu einem jeweiligen sendenden Anwender gehört, und letzterer wird anhand der gespeicherten Sprachprobe und des zugehörigen Referenzbildes identifiziert, wozu eine Bildanalyse und
20 eine Sprachanalyse ausgeführt werden.

Bei der Bildanalyse werden bevorzugt charakteristische Gesichtszüge eines jeweiligen Anwenders mit einem zugehörigen Referenzbild verglichen.

25 Bevorzugt wird ein jeweiliger Anwender vor Ausführung des Verfahrensschritts c) als berechtigt identifiziert, die gespeicherten Referenzdaten zur Editierung des von ihm aufgenommenen Bildes und gespeicherte Sprachänderungs-
30 algorithmen zur Änderung der Stimmlage seiner Sprache zu verwenden.

Das erfindungsgemäße Verfahren umfasst die anwenderspezifische Optimierung von Bewegtbildern, zweidimensionalen und dreidimensionalen Bilddaten.

5

Vorzugsweise wird das zu übertragende Bild, das gegebenenfalls anwenderspezifisch editiert wurde, vor der Übertragung dem jeweiligen Anwender angezeigt.

10

Eine zur Durchführung des obigen Verfahrens zur darstellungsoptimierten Übertragung von Video- und Audiodaten eingerichtete Vorrichtung ist erfindungsgemäß dadurch gekennzeichnet, dass die Vorrichtung aufweist:

- Videoeingabe- und Ausgabemittel;
- 15 - Audioaus/eingabemittel;
- Sende- und Empfangsmittel;
- eine Schnittstelle zu einem Übertragungskanal;
- eine Eingabevorrichtung zur Eingabe von Steuer- und Befehlssignalen;
- 20 - eine Speichervorrichtung zur Speicherung von Anwender- und Systemprogrammen sowie von Referenzbilddaten und -audiodaten, wobei

die vorgenannten Mittel und Komponenten mit einer Prozessoreinheit in funktioneller Verbindung stehen, die
25 dazu eingerichtet ist, die Verfahrensschritte in Interaktion mit den genannten Mitteln und Komponenten auszuführen.

Weitere vorteilhafte Merkmale ergeben sich aus der nachfolgenden Beschreibung bevorzugter Ausführungsbeispiele, die sich auf die beiliegende Zeichnung bezieht.

5 Zeichnung

Figur 1 ist ein Blockdiagramm einer erfindungsgemäßen Vorrichtung zum Senden/Empfang audiovisueller Informationen mit anwenderoptimierten Inhalt.

10

Figur 2 ist ein Blockdiagramm über Details der Funktionsgruppe "Speicher" gemäß Figur 1.

15

Figur 3 ist ein Blockdiagramm von Funktionen bei der Anwendung der Vorrichtung der Figuren 1 und 2 ohne anwenderkontrollierte Optimierung.

20

Figur 4 ist ein Blockdiagramm von Funktionen mit Anwendung einer anwenderkontrollierte Optimierung.

Figur 5 ist ein Funktionsflussdiagramm von Verfahrensschritten, wenn keine anwenderkontrollierte Optimierung ausgeführt wird, und

25

Figur 6 ist ein Funktionsflussdiagramm von Verfahrensschritten, wenn eine anwenderkontrollierte Optimierung ausgeführt wird.

30

Ausführungsbeispiele

Die Erfindung geht von dem Vorhandensein audiovisueller Kommunikationsmedien aus. Allgemeine Merkmale audiovisueller Kommunikationsmedien sind Mikrophon und Lautsprecher, Videokamera und Bildschirm, eine Steuerungseinheit, eine Verarbeitungseinheit abgehend zur Verarbeitung von Audio- und Videosignalen, einer Verarbeitungseinheit kommend zur Verarbeitung von Audio- und Videosignalen und einer Kompressionseinheit zur optimalen Ausnützung der zur Verfügung stehenden Leitungsbandbreite, z. B. über analoge und digitale Telefonnetze, paketgesteuerte Kommunikation über das Internet, interne Computernetze usw.

In der nachfolgenden Beschreibung werden die Begriffe "Teilnehmer" und "Anwender" synonym verwendet. Die vorliegende Erfindung erweitert die oben genannten Merkmale audiovisueller Kommunikationsmedien um eine Funktion, die nachfolgend "optimierende Bildverarbeitung" (kurz: OBV) genannt wird. Die erfindungsgemäße Funktion OBV kann sowohl in Videobildtelefonen, Videokonferenzsystemen oder auf Internetprotokoll basierten Systemen angewendet werden. Es soll erwähnt werden, dass sich die Beschreibung hauptsächlich mit der anwenderkontrollierten OBV befasst, und die Sprachoptimierung nicht so detailliert behandelt wird.

Die erfindungsgemäße OBV gibt dem Teilnehmer die Möglichkeit, 1-n optimierte Abbilder seiner gewünschten Erscheinung als Referenzbildinformation in einem oder auch

08.10.99

12

9

mehreren Speichermedien abzuspeichern. Dabei können auch Speichermedien unterschiedlicher physikalischer Art zum Einsatz kommen.

5 Nun wird Bezug auf Figur 1 genommen, die ein Blockdiagramm einer Vorrichtung zeigt, die zur Ausführung des erfindungsgemäßen Verfahrens, d. h. zum Senden/-Empfang audiovisueller Informationen mit anwenderkontrollierter Optimierung eingerichtet ist.

10

Gemäß Figur 1 steht eine zentrale Prozessoreinheit 10 in funktioneller Verbindung mit einer Bildeingabeeinheit 11, z. B. einer Videokamera, mit einer Anzeigevorrichtung 12, z. B. einem PC-Monitor oder einer Flüssigkristallanzeige, mit einer Audioaus/eingabeeinrichtung 13, mit einer audiovisuellen Kommunikationseinheit 14, 15, 16, die aus einer Sendeeinheit 14, einer Empfangseinheit 15 und einer Schnittstelle 16 zu einem Telefonnetz, Funknetz, Mobilnetz oder Datennetz besteht, mit einer Schnittstelle 17 zu einem übergeordneten Speichermedium oder einem Rechner, wie z. B. einem Personalcomputer, mit einer Eingabevorrichtung 18, z. B. einer Telefontastatur oder einer separaten Tastatur und mit einem Speicher 20, in dem zumindest anwenderspezifische Programme 21 und Referenzbilder 22 und gegebenenfalls Audioreferenzdaten gespeichert sind.

Figur 2 zeigt, dass die Speichervorrichtung 20 mehrere, auch physikalisch unterschiedliche Speichermedien aufweisen kann. Die optimierten Referenzinformationen, d. h. die Referenzbildinformation und die Referenzaudio-

information, werden zum einen über die Bildeingabeeinheit 11, die eine Videokamera, ein Rekorder, etc., sein kann, die Audioeingabeeinheit 13, oder von einer übergeordneten Einheit über die Schnittstelle 17 in die Speichervorrichtung 20 übertragen. Dabei ist es von untergeordneter Rolle, ob die Anwendungen und das Bildmaterial z. B. in einem ROM-Speicher, einem RAM-Speicher oder einem Massenspeicher wie z. B. in einer Festplatte, in Flashcards oder ähnlichen Medien zur Verfügung gestellt werden (Figur 2). Wie erwähnt, kann der Speicher 20 auch Referenzaudioinformation enthalten, die zur anwenderkontrollierten Optimierung seiner Sprache verwendet werden kann. Durch die Eingabevorrichtung 18 steuert der Anwender die optimierende Bildverarbeitung, um nach seiner Wahl das zu sendende Bild oder seine Sprache zu optimieren. Die Eingabevorrichtung 17 kann z. B. eine Telefontastatur, ein separat angeschlossenes Tastaturfeld 18, eine Computermouse, ein Lichtstift, ein "graphics tablet" etc. sein.

In Figur 2 sind Einzelheiten der Speichervorrichtung 20 sowie der darin gespeicherten anwenderspezifischen Informationsinhalte 21 und der Referenzbilder 22 gezeigt. Wie dargestellt, kann die Speichervorrichtung 20, die bei der Erfindung verwendbar ist, einen ROM-Speicher, RAM-Speicher, eine Festplatte, eine Wechselplatte, eine Floppy Disk, eine Flashcard und andere geeignete Speichermedien aufweisen. Sie kann auch eine Kombination solcher Speichermedien enthalten. In dem Inhaltsblock 21, in dem anwenderspezifische Programme und Informationen gespeichert sind, befindet sich ein Block 210 für die

Benutzererkennung, ein Block 211 für die technische Bildoptimierung mit verschiedenen Algorithmen 1, 2, ...N, ein Block 212 für die kosmetische Erscheinungsoptimierung, wieder mit verschiedenen Bildbearbeitungsalgorithmen 1, 2, ...N, ein Block 213 zur Hintergrundbearbeitung, ein Block 214 zur Kopfbearbeitung und ein Block 215 zur Körperbearbeitung. Es muss hier bemerkt werden, dass die im anwenderspezifischen Block 21 enthaltenen und oben aufgeführten Einzelblöcke 211-215 nicht unbedingt vollständig aufgezählt wurden und dass Speicherinhalte zur Spracherkennung und weitere Inhalte, die hier nicht wesentlich sind, enthalten sein können.

Der in Figur 2 dargestellte Block 22, der das Bildmaterial, d. h. die Referenzbildinformation betrifft, enthält einen Block 221 mit Referenzhintergrundbildern 1, ...N, einen Block 222 mit Referenzbildern 1, 2, ...N des Kopfs und einem Block 223 mit Referenzbildern 1, 2, ...N des Körpers des Anwenders. Es muss hier erwähnt werden, dass der Begriff "Körper" die Körperteile außerdem Kopf umfassen.

Für eine audiovisuelle Kommunikationsvorrichtung gemäß der Erfindung, die für mehrere Teilnehmer berechtigt ist, sind die Blöcke 221-223 entsprechend mehrfach vorhanden.

In Figur 3 ist in Form eines Funktionsblockschaltbildes eine Funktionsweise ohne senderseitige optimierende Bildbearbeitung, d. h. ohne die erfindungsgemäße OBV. Die oben genannte OBV, die in einem Funktionsblock 40 dargestellt ist, ist deaktiviert, und das Videosignal von

der Videosignalquelle geht direkt zu einem Kodierungsblock 41.

5 Dagegen zeigt Figur 4 in Form eines Funktionsblockschaltbildes den Fall, wo eine optimierende Bildverarbeitung ausgeführt wird.

10 Vor der Sendung wird das von einer Videosignalquelle 11 erhaltene Videosignal zur Vorbereitung der Optimierung im Sinne der Erfindung durch einen Inhaltserkennungsblock 30 in einzelne Ebenen 31-33 (englisch: "layers") aufgeschlüsselt. Die Ebenen 31-33 sind hier nur zum Zwecke des besseren Verständnisses als Ebene 1 (Hintergrund), Ebene 2 (Körper) und Ebene 3 (Kopf) bezeichnet. Diese Ebenen 15 entsprechen den im Speicherbereich 22 gespeicherten Schichten Hintergrund 221, Kopf 222 und Körper 223 (vergleiche Figur 2). Diese Begriffe Hintergrund, Körper, Kopf dienen somit nur zur Veranschaulichung der Erfindung. Je nach Größe des bereitgestellten Speichers 20 können weitere Ebenen zur weiteren Detaillierung definiert sein. Ebenfalls zur Veranschaulichung des Verfahrens wird eine weitere, gestrichelt eingezeichnete, Spezialebene 34 definiert, die wahlweise zugeschaltet werden kann. Die beschriebenen Ebenen 31-34 werden aus 25 dem Videosignalstrom von der Videosignalquelle 11, z. B. von der Videokamera, mit Bildverarbeitungsalgorithmen im Inhaltserkennungsblock 30 herausgelöst und separat in logischen Speicherebenen verwaltet. Findet eine audiovisuelle Kommunikation statt, werden je nach Einstellung 30 des Anwenders alle oder einzelne Ebenen durch Bildinformation aus dem Speicher 22 ersetzt. Im Beispiel von

Figur 4 werden die Ebenen 31 und 32, betreffend Hintergrund und Referenzkopf, ersetzt durch ein Referenzhintergrundbild 2 und ein Kopfbild 3 aus den Speicherbereichen 221 und 222. Die Steuerung der eingesetzten Bildinformation wird durch Referenzmarken sichergestellt.

Als Beispiel soll das Gesicht dienen. Die erkannten Referenzmarken werden zur Steuerung der einzusetzenden Bildinformation der Ebene "Kopf" verwendet. Bewegt der Anwender den Kopf, wenn er z. B. zustimmend nickt, so führt das optimierte Bild die gleiche Bewegung aus. Ist die Ebene "Körper" aktiviert, wird die aktuelle Ebene "Körper" durch eines der gespeicherten Körperreferenzbilder im Bereich 223 ersetzt. Alle Ebenen zusammengesetzt bilden das zu sendende optimierte Videobild.

Es muss hier bemerkt werden, dass die Bezeichnung "Ebene" nicht bedeutet, dass mit dem erfindungsgemäßen Verfahren verarbeitete bzw. editierte Bildinformation nur zweidimensional sein kann. Statt dessen können auch dreidimensionale Bildinformationen verarbeitet werden.

Das neu zusammengesetzte Videobild wird jetzt technisch hinsichtlich Farbwerte, Kontraste, Helligkeit nach voreingestellten oder voreinstellbaren Parametern optimiert (siehe Block 42 in Figur 4).

In einer weiteren Funktionseinheit 43, die kosmetische Bildoptimierung genannt ist, werden bevorzugt im Gesichtsbereich kosmetische Verbesserungen vorgenommen. Dazu gehört das Aufhellen der Augen und die Augenwinkel-

korrektur, Aufhellen des Zahnbereichs, Schattenaufhellung (z. B. bei Bartwuchs und bei großflächigen Pigmentveränderungen der Haut) und das Wegretuschieren von nicht gewünschten kleinen Pigmentstörungen, wie z. B. Warzen.

5

Das so optimierte Videosignal wird an die Kodierungseinheit 41 weitergegeben und schließlich an den oder die Empfänger gesendet.

- 10 Nachstehend werden bezogen auf die Figuren 5 und 6 Kommunikationsschritte unter Einsatz des erfindungsgemäßen Verfahrens beschrieben. Zunächst wird eine Kommunikationsverbindung von außerhalb beschrieben. Die audiovisuelle Kommunikationsvorrichtung (Figur 1) erhält
- 15 ein Signal von der Gegenseite zum Aufbau einer Verbindung. Dem Stand der Technik entsprechende Kommunikationssysteme übermitteln eine sog. Anruferkennung (englisch: "Caller Recognition"). Bei eingehendem Anruf (Schritt 51) wird die Anruferkennung überprüft (Schritt
- 20 53). Dies erfolgt durch Vergleich der übermittelten Anruferkennung mit einem Teilnehmerverzeichnis 52 (Telefonbuch mit Teilnehmeradressen), das im Speicher gespeichert ist. Den Einträgen des Teilnehmerverzeichnis 52 können bestimmte OBV-Konfigurationen zugeordnet werden. Im vorgestellten Beispiel wird davon
- 25 ausgegangen, dass der Anrufer entweder keinen Eintrag im Teilnehmerverzeichnis 52 hat oder dass dem Eintrag im Teilnehmerverzeichnis 52 die OBV-Konfiguration "Büro" zugeordnet wurde. In diesen Fällen wird das Signal der
- 30 Videokamera 11 an den Inhaltserkennungsfunktionsblock 30 weitergeleitet, der Schritt 57 berechnet die Ebenen

(31-34 in Figur 4), die im Schritt 58 mit den Referenzinformationen der OBV-Konfiguration "Büro" in Echtzeit zusammengesetzt werden. Danach wird die bereits erwähnte kosmetische Bildoptimierung 43 und die
5 technische Bildoptimierung 42 ausgeführt. Nach der technischen Bildoptimierung 42 entsteht im Block 40 das optimierte Videobild nach den voreingestellten Parameterwerten in Zuordnung zum Teilnehmereintrag im Verzeichnis 52. Die Signale, d. h. die optimierten Bilder und
10 Audiosignale werden anschließend je nach Kommunikationsprotokoll im Funktionsblock 41 kodiert und übertragen.

Nun wird bezogen auf Figur 5 der Fall betrachtet, dass a) der gewünschte Teilnehmer dem Teilnehmerverzeichnis 52
15 bekannt ist (Ausgang "JA" des Vergleichsschritts 53) und dass b) dem Teilnehmer die OBV-Konfiguration "Privat" zugeordnet ist (Schritt 55). Dann wird das unbearbeitete Videosignal, d. h. ohne OBV, der Kodierungseinheit 41 übermittelt.

20 Anschließend wird Bezug auf Figur 6 genommen, die einen Verbindungsaufbau vom vorliegenden Anwender darstellt. Wenn der vorliegende Anwender eine Kommunikation aufbauen möchte, hat er die Möglichkeit, einen Teilnehmer aus dem
25 Teilnehmerverzeichnis auszuwählen (Schritt 62) oder den entsprechenden Teilnehmercode manuell einzugeben (Schritt 61). Gesetzt der Fall, der gewünschte Teilnehmer wird aus dem Teilnehmerverzeichnis 62 gewählt, lautet die Antwort auf die im Schritt 64 gestellte Frage: Code ID bekannt?
30 "Ja". Dann wird nach Maßgabe des OBV-Parameters im Teilnehmerverzeichnis 62 die OBV-Konfiguration "Freizeit"

aktiviert (Schritt 66). In diesem Fall wünscht der Anwender nur eine kosmetische Bildoptimierung im Funktionsblock 43 und die technische Bildoptimierung im Funktionsblock 42. Das hierdurch veränderte Videosignal wird
5 sodann nach den Erfordernissen des Kommunikationsprotokolls im Funktionsblock 41 kodiert.

Anschließend wird ein Beispiel beschrieben, wo der Anwender den erforderlichen Verbindungscode im Schritt 61
10 manuell über die Eingabeeinheit 18 (Figur 1) eingibt. Die im Abfrageschritt 64 ausgeführte Überprüfung der manuell eingegebenen Teilnehmernummer oder -adresse ergibt im Beispiel eine negative Antwort bei der Abfrage 64. Daraufhin aktiviert die OBV die Konfiguration "Büro".
15 Daraufhin wird die komplette Bildbearbeitung, d. h. Optimierung der Bildinformation und gegebenenfalls der Audioinformation ausgeführt. Die Ausführung der Optimierung der Audioinformation ist in den Figuren 5 und 6 nicht dargestellt. Das editierte Signal wird dann an die
20 Kodierungseinheit 41 weitergegeben.

Abgesehen von dieser automatischen Auswahl zwischen der vollständigen OBV-Funktion und der "rudimentären Editierung" der Bildinformation durch die kosmetische
25 Bildoptimierung 43 und die technische Bildoptimierung 42, hat der Anwender jederzeit die Möglichkeit, über die Eingabeeinheit 18 die OBV-Konfiguration während einer Kommunikation zu aktivieren oder zu deaktivieren.

30 Das vorbeschriebene erfindungsgemäße Verfahren kann auch dazu verwendet werden, um z. B. Gesichter anderer

Menschen oder gänzlich künstlich erstellte Charaktere zu übertragen. Dabei kann der Anwender als "Animateur" auftreten. Die durch die Erfindung ermöglichte Durchführung solcher "Animationen" ist wahrscheinlich kommerziellen Anwendung vorbehalten. Hier wird z. B. vom Wunsch von Unternehmen ausgegangen, einen bestimmten Charakter als Marke zu etablieren. Zur Veranschaulichung sei der Charakter "Herr Kaiser" eines Versicherungsunternehmens und außerdem der Charakter "Micky Maus" der Walt Disney Corporation angeführt. Eine kommerzielle Anwendung ist hierbei die Bereitstellung einer erfindungsgemäßen OBV für Telefonagenturen. Unternehmen beauftragen Telefonagenturen, um Kundenanfragen und -wünsche zu beantworten oder aufzunehmen. Kunden des oben als Beispiel angeführten Versicherungsunternehmens kommunizieren auf diese Weise audiovisuell mit dem durch OBV etablierten Sympathieträger des Unternehmens. Kinder und Jugendliche können über sympathische Comicfiguren, wie z. B. Micky Maus, durch das Unternehmen Informationen über spezielle Aktionen und Tips zur Freizeitgestaltung erfahren.

Um einen Missbrauch des vorbeschriebenen Verfahrens zu unterbinden, stehen Algorithmen zur Gesichtserkennung zur Verfügung. Die Gesichtserkennung ist Bestandteil der Steuerung der oben beschriebenen Ebene "Kopf". Sofern der Anwender nicht die gleichen charakteristischen Gesichtsm征kmale wie das abgespeicherte Kopfbild hat, wird das Videosignal uneditiert an die Kodierungseinheit zur Sendung weitergegeben.

08.10.99

21

18

Weiterhin enthält das vorbeschriebene Verfahren zur Auswahl von berechtigten Teilnehmern Spracherkennungsalgorithmen, welche vor allem in kommerziellen Anwendungen aktiviert sind. Ein aktueller Teilnehmer spricht eine
5 Kennphrase, die mit gespeicherten Sprachproben verglichen wird. Ergibt sich eine Übereinstimmung, kann zusätzlich noch eine Korrelation mit der erwähnten Bildererkennung anhand eines gespeicherten Kopfbildes hergestellt werden. Somit ist ein Missbrauch der Erfindung durch Personen-
10 verwechslung oder Vortäuschung von anderen Personen ausgeschlossen.

00.10.99

VOXAR AG
VOX/P991

**Verfahren und Vorrichtung zur darstellungs-
optimierten Übertragung von Video- und Audiodaten**

5

PATENTANSPRÜCHE

- 10 1. Verfahren zur darstellungsoptimierten Über-
tragung von Videodaten und/oder Audiodaten, insbesondere
beim Bildfernsprechen, bei dem von einer Video- und
Audioquelle kommende Bild- und Audiodaten vor ihrer
Übertragung zu einem Empfänger anhand von
15 Referenzinformation entsprechend wenigstens einem
vorbestimmten oder vorbestimmbaren Kriterium verändert
werden, **dadurch gekennzeichnet, dass** das Verfahren
wenigstens folgende Schritte aufweist:
- 20 a) Referenzbildinformation wird zuvor erstellt und
gespeichert;
- b) Parameter, die Bildinformation betreffen, werden
zuvor definiert, gespeichert und der in Schritt a)
gespeicherten Referenzbildinformation zugeordnet;
- 25 c) ein von einer Videoquelle kommendes Bild wird
hinsichtlich eines oder mehrerer, ausgewählter
(auswählbarer) Bildinformationsparameter von den in
Schritt b) definierten und gespeicherten Parameter
extrahiert;
- 30 d) das Bild von Schritt c) wird basierend auf dem oder
den ausgewählten Bildinformationsparameter(n) anhand der
zugeordneten Referenzbildinformation editiert; und

08.10.99

2

e) das in Schritt d) editierte Bild wird zu einem oder mehreren Empfängern übertragen.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet,**
5 **dass** die Schritte d) und e) an einem zentralen Ort ausgeführt werden, der vom sendenden Anwender entfernt ist, wobei das Bild, das zugeordnete Referenzbild und der oder die Bildinformationsparameter vom Ort des sendenden Anwenders zum zentralen Ort übertragen werden.

10 3. Verfahren nach Anspruch 1, **dadurch gekennzeichnet,**
dass die Schritte a) bis e) am Ort eines sendenden Anwenders ausgeführt werden.

15 4. Verfahren nach einem der vorangehenden Ansprüche, **dadurch gekennzeichnet, dass** es einen zusätzlichen Schritt f) aufweist, mit dem zuvor gespeicherte Sprachproben berechtigter Anwender mit einer gesprochenen Kennphrase eines aktuell sendenden Anwenders verglichen
20 werden und bei positivem Vergleichsergebnis eine Sende-freigabe für diesen Anwender erfolgt.

5. Verfahren nach Anspruch 4, **dadurch gekennzeichnet,**
dass in Schritt f) außerdem die gespeicherte Sprachprobe
25 einem Referenzbild, das zu einem jeweils sendenden Anwender gehört, oder als zu ihm gehörig ausgewählt wurde, zugeordnet wird, und der jeweilige sendende Anwender anhand der gespeicherten Sprachprobe und des zugehörigen Referenzbildes identifiziert wird.

30

08.10.99

24

3

6. Verfahren nach Anspruch 4 oder 5, **dadurch gekennzeichnet, dass** in Schritt f) eine Sprachanalyse der gesprochenen Kennphrase und eine Bildanalyse des von der Videoquelle kommenden Bildes ausgeführt werden.

5

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, dass** bei der Bildanalyse charakteristische Gesichtszüge eines jeweiligen sendenden Anwenders mit einem zugehörigen oder ausgewählten Referenzbild verglichen werden.

10

8. Verfahren nach einem der Ansprüche 5-7, **dadurch gekennzeichnet, dass** der jeweilige sendende Anwender vor Ausführung des Schrittes c) als berechtigt identifiziert wird, gespeicherte Referenzinformation zur audiovisuellen

15

9. Verfahren nach einem der vorangehenden Ansprüche, **dadurch gekennzeichnet, dass** die zu verarbeitenden Bild-
daten, die analysierte und editierte Bildinformation
20 Bewegtbilder, zweidimensionale und dreidimensionale Bild-
information umfassen.

20

10. Verfahren nach einem der vorangehenden Ansprüche, **dadurch gekennzeichnet, dass** das zu übertragende Bild vor
25 der Übertragung dem sendenden Anwender angezeigt wird.

25

11. Verfahren nach Anspruch 10, **dadurch gekennzeichnet, dass** die Anzeige eine interaktive Benutzeroberfläche aufweist, die auch die auswählbaren und/oder ausgewählten
30 Referenzbilder sowie die auswählbaren oder ausgewählten
Bildinformationsparameter anzeigt.

30

12. Verwendung des Verfahrens nach einem der Ansprüche 1-11 zur darstellungsoptimierten Übertragung von Video- und/oder Audiodaten beim Bildfernsprechen.

5

13. Verwendung des Verfahrens nach einem der Ansprüche 1-11 zur darstellungsoptimierten Übertragung von Video- und/oder Audiodaten bei Videokonferenzen.

10

14. Verwendung des Verfahrens nach einem der Ansprüche 1-11 zur darstellungsoptimierten Übertragung von Video- und/oder Audiodaten über Computernetze.

15

15. Vorrichtung zur darstellungsoptimierten Übertragung von Video- und/oder Audiodaten, insbesondere zur Durchführung des Verfahrens nach einem der Ansprüche 1-14, **dadurch gekennzeichnet, dass** die Vorrichtung aufweist:

20

- Videoeingabe- und Ausgabemittel (11, 12);

- Audioaus/eingabemittel (13);

- Sende- und Empfangsmittel (14, 15);

- eine Schnittstelle (16) zu einem Übertragungskanal;

- eine Eingabevorrichtung (18) zur Eingabe von Steuer- und Befehlssignalen;

25

- eine Speichervorrichtung (20) zur Speicherung von Anwender- und Systemprogrammen sowie von Referenzbilddaten und -audiodaten, wobei

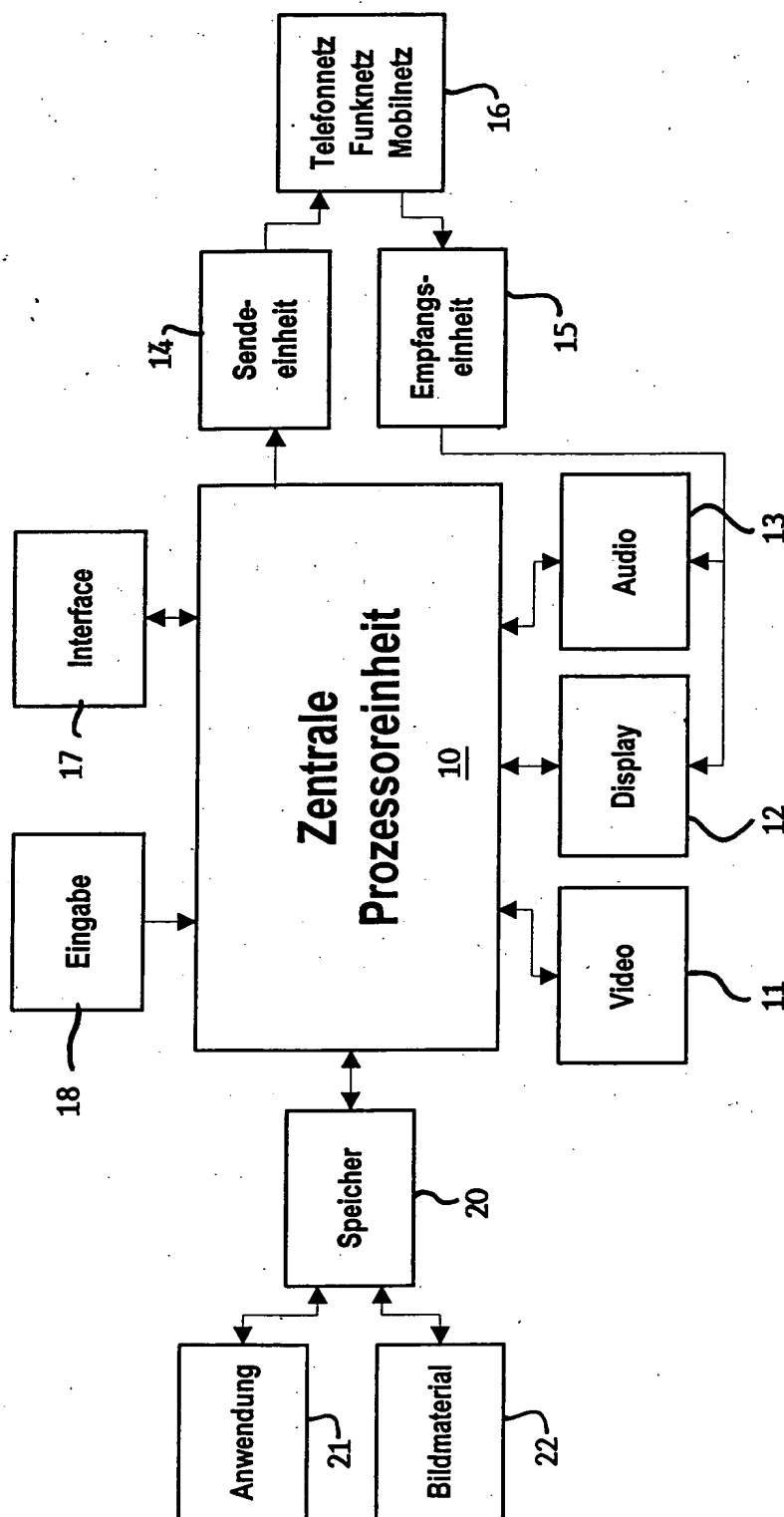
30

die vorgenannten Mittel und Komponenten mit einer Prozessoreinheit (10) in funktioneller Verbindung stehen, die dazu eingerichtet ist, die Verfahrensschritte in

Interaktion mit den genannten Mitteln und Komponenten auszuführen.

16. Vorrichtung nach Anspruch 15, **dadurch gekennzeichnet,** dass sie weiterhin eine mit der Prozessor-
5 **zeichnet, dass** sie weiterhin eine mit der Prozessor-
einheit (10) in Verbindung stehende Schnittstelle (17)
aufweist zur Verbindung mit einer übergeordneten Ver-
waltungseinheit und/oder einem übergeordneten Speicher-
medium, z. B. mit einem "Personal.Computer".

FIG. 1



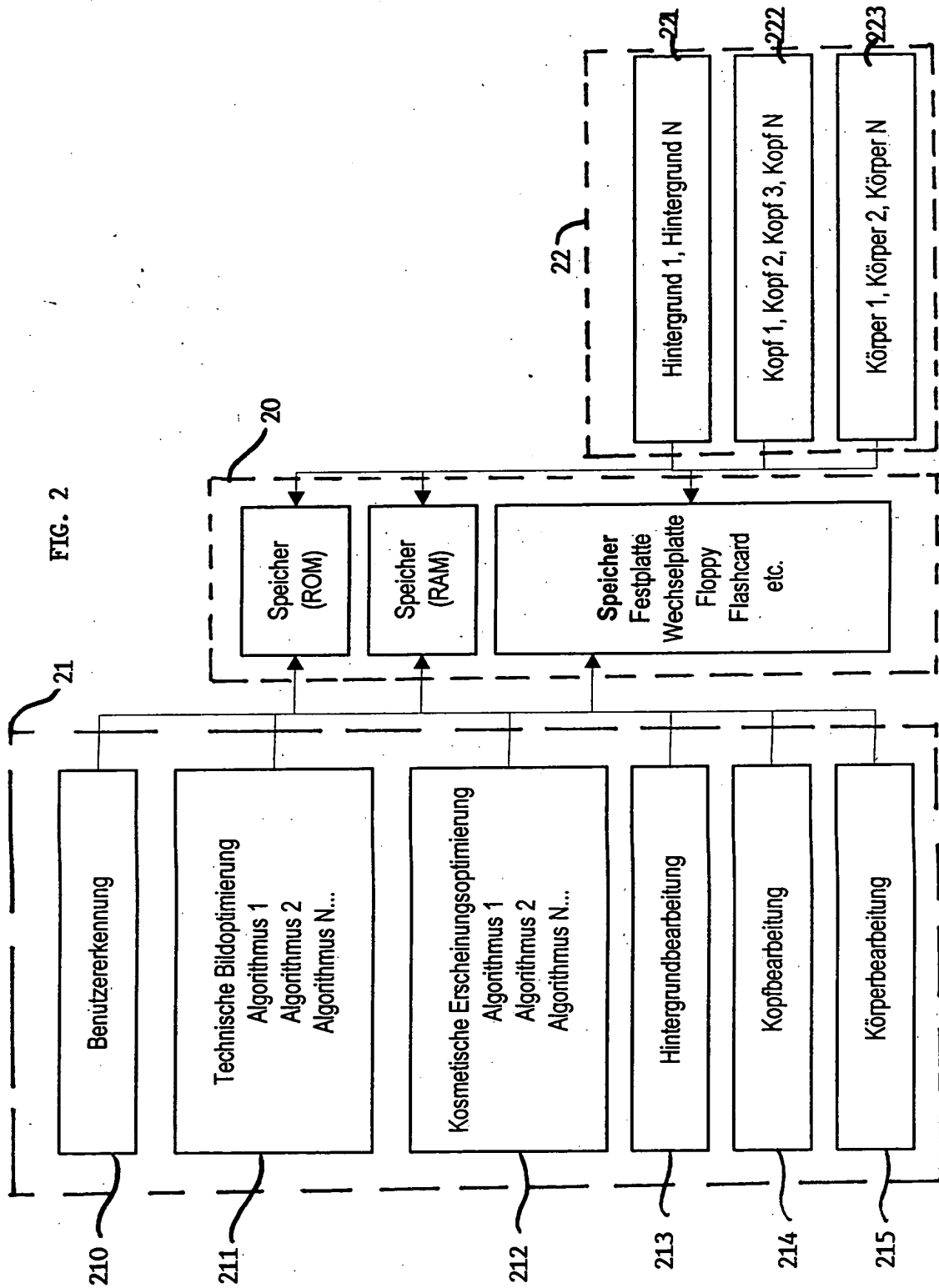


FIG. 3

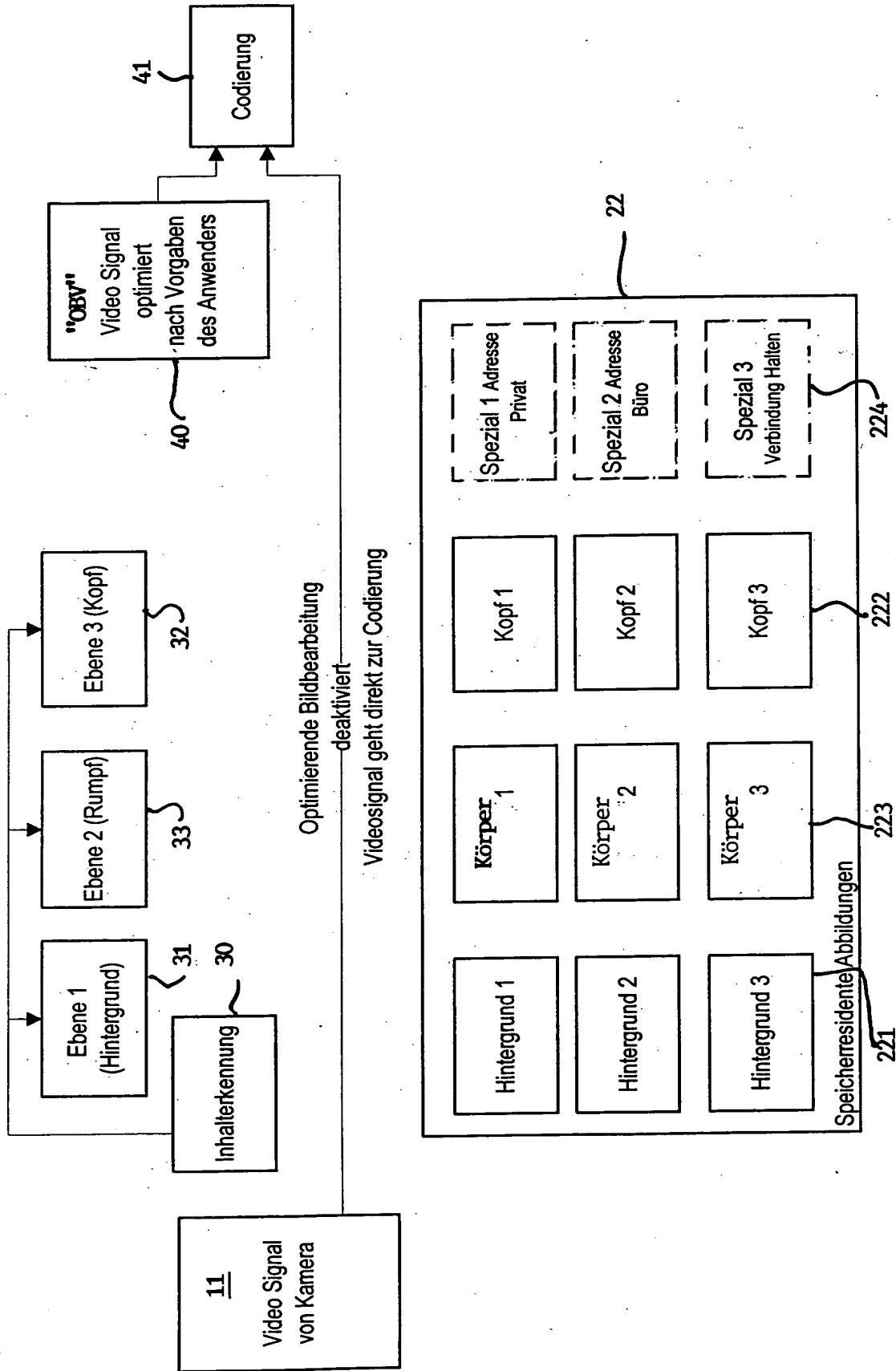
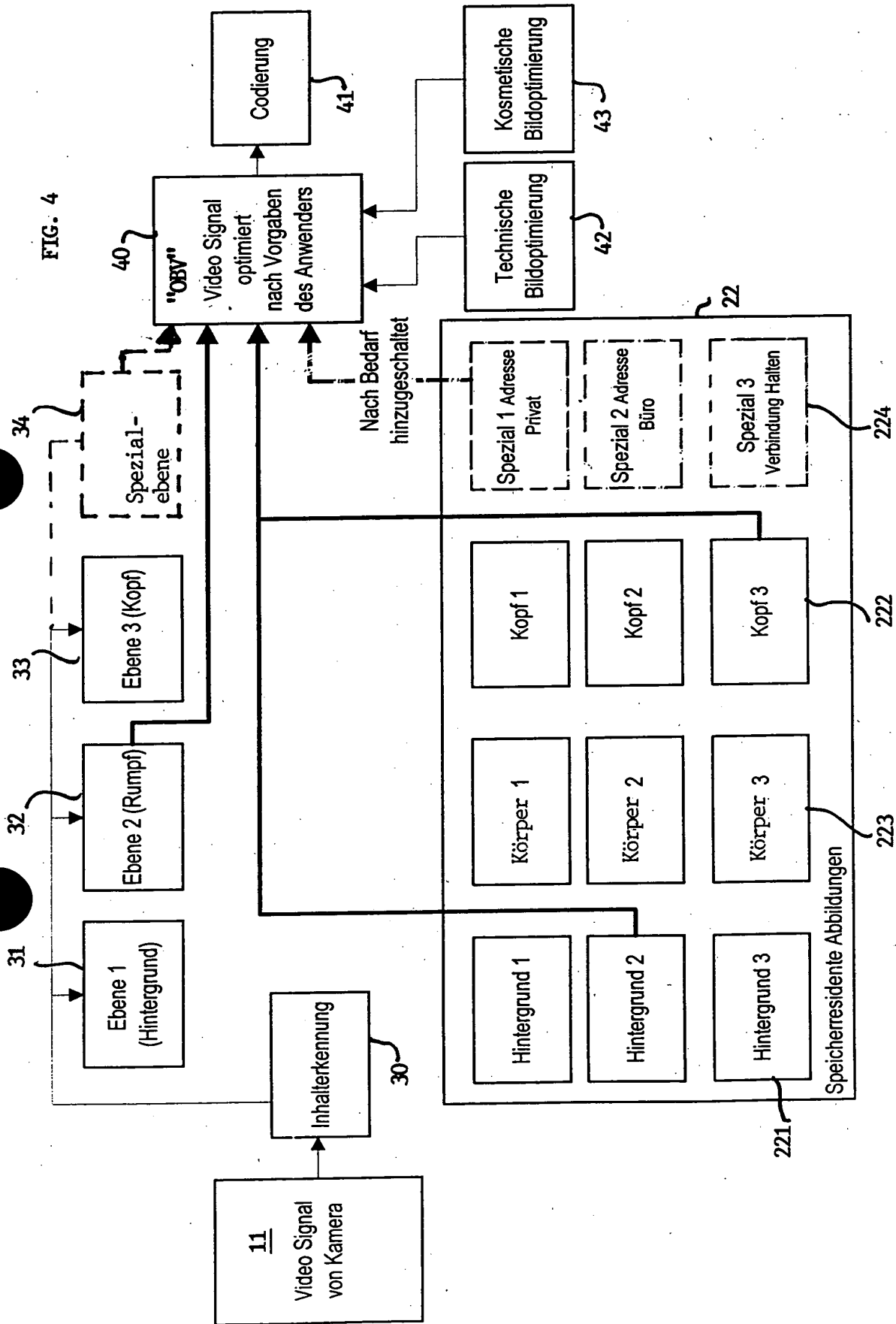


FIG. 4



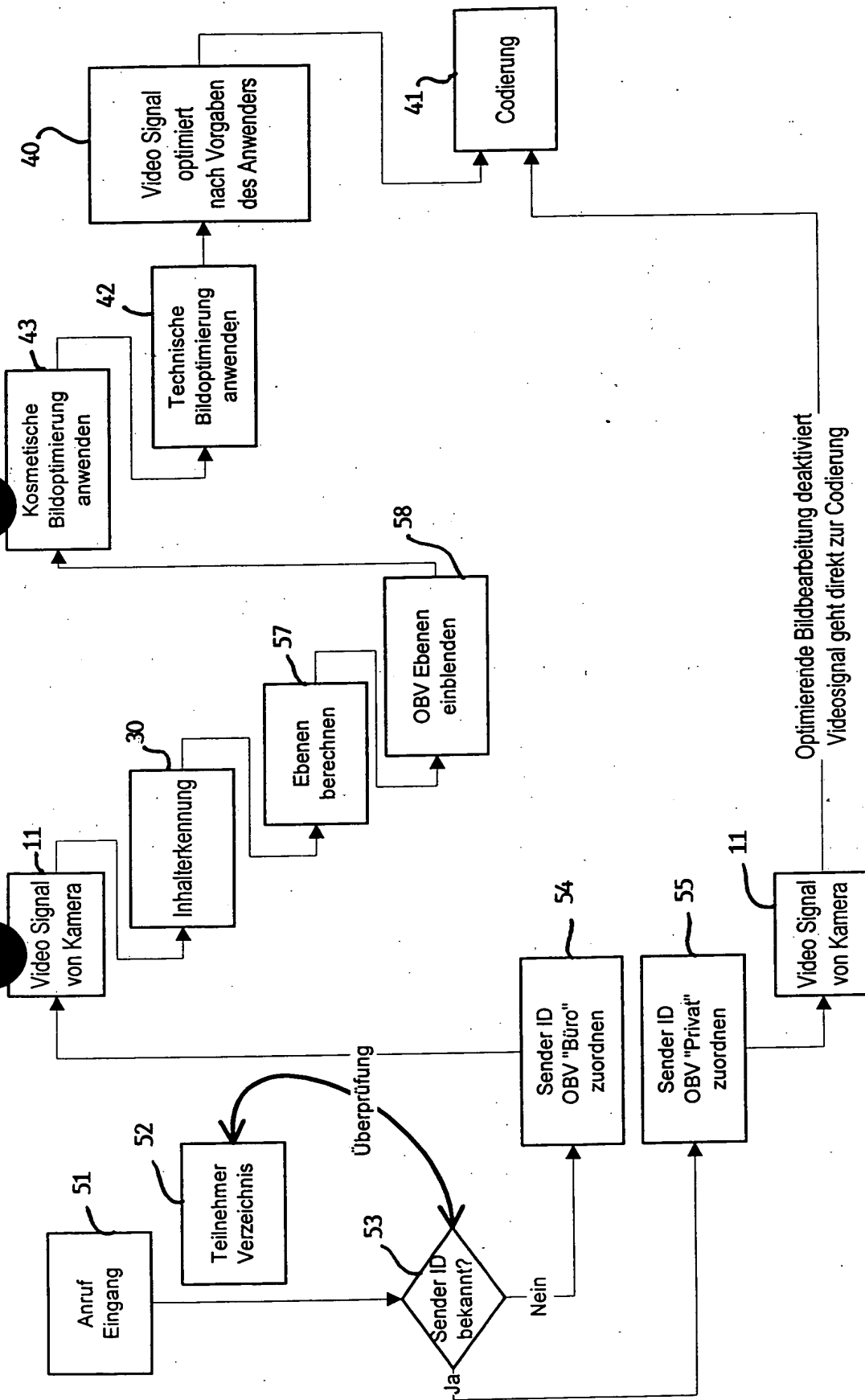
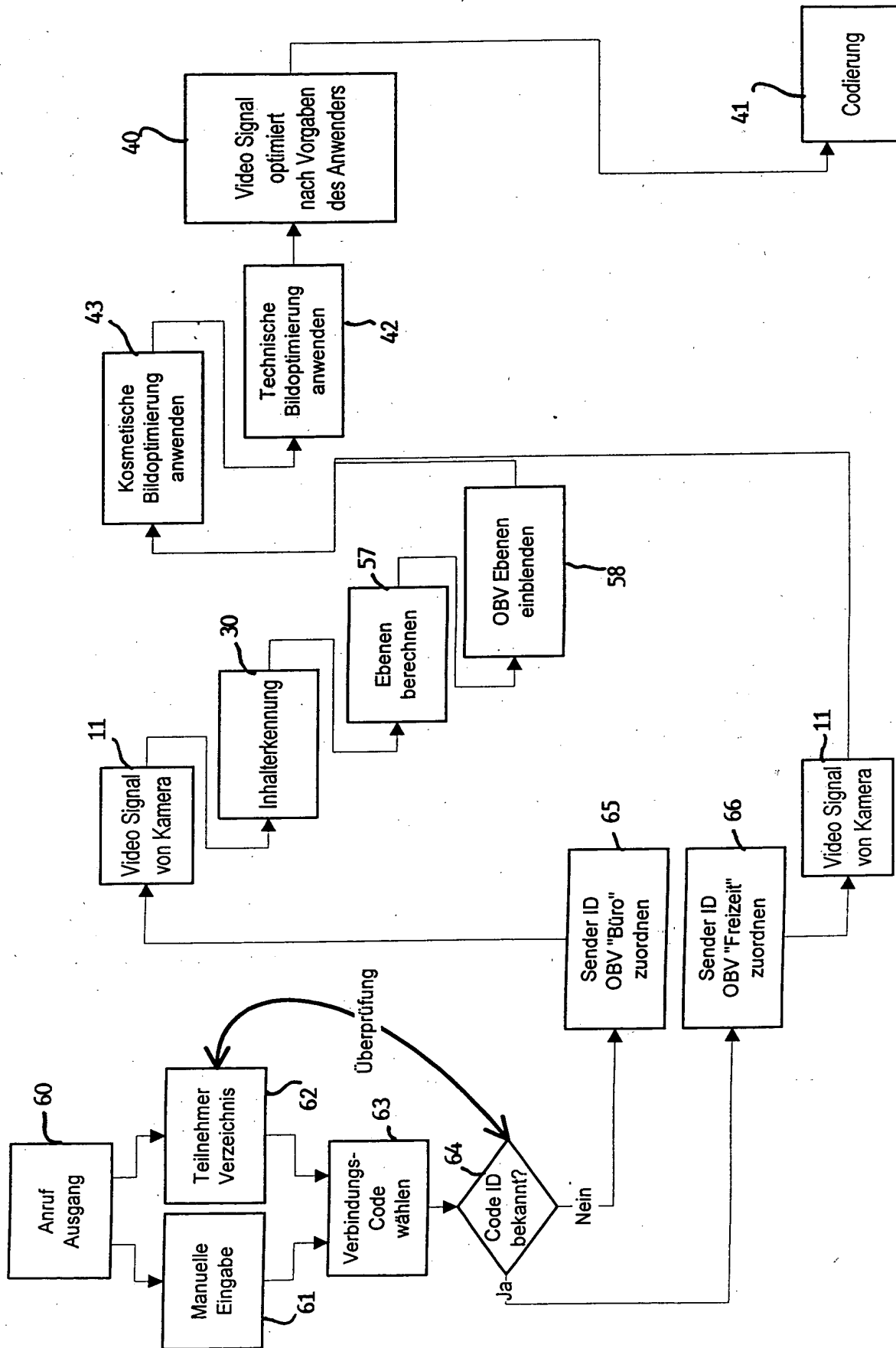


FIG. 5

FIG. 6



08.10.99

3

VOXAR AG
VOX/P991

ZUSAMMENFASSUNG

Die Erfindung betrifft ein Verfahren und eine Vorrichtung für die ästhetische und inhaltliche Veränderung und Optimierung des Videoinhalts und Audioinhalts zwischen zwei oder mehr Teilnehmern einer audiovisuellen Kommunikationsverbindung, wie Bildtelefon über Festanschluss, Bildtelefon über Short-Range und Long-Range Funkverbindung sowie andere audiovisuelle Kommunikationsmittel, z. B. über Personal Computer, die in einem Netz verbunden sind. Von einer Videosignalquelle (11) kommende Bilder eines Teilnehmers werden einer Inhaltserkennung (30) unterworfen und in verschiedene Ebenen (31-34) unterteilt, wie z. B. in Hintergrund, Körper, Kopf. Dabei kann nach Wahl des Anwenders eine oder mehrere dieser Ebenen durch ein in einem Speicher abgelegtes Referenzbild oder mehrere Referenzbilder ersetzt werden. Dies geschieht in einer "optimierenden Bildverarbeitung" (40). Zusätzlich kann eine technische und kosmetische Bildoptimierung (42, 43) ausgeführt werden. Das erfindungsgemäße Verfahren enthält zudem Algorithmen zur Erkennung von Gesichtszügen eines Teilnehmers sowie zur Spracherkennung, so dass durch Vergleich des abgespeicherten Referenzgesichtsbildes mit dem Gesicht des Teilnehmers sowie einer Kennungsphrase mit abgespeicherten Sprachproben ermittelt werden kann, ob ein aktueller Anwender berechtigt ist oder nicht (Figur 4).

THIS PAGE BLANK (USPTO)